

A Variational Framework for the Simultaneous Segmentation and Object Behavior Classification of Image Sequences

Laura Gui¹, Jean-Philippe Thiran¹, and Nikos Paragios²

¹ Signal Processing Institute (ITS), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

² Laboratoire MAS - Ecole Centrale de Paris, Châtenay-Malabry, France

Abstract. In this paper, we advance the state of the art in variational image segmentation through the fusion of bottom-up segmentation and top-down classification of object behavior over an image sequence. Such an approach is beneficial for both tasks and is carried out through a joint optimization, which enables the two tasks to cooperate, such that knowledge relevant to each can aid in the resolution of the other, thereby enhancing the final result. In particular, classification offers dynamic probabilistic priors to guide segmentation, while segmentation supplies its results to classification, ensuring that they are consistent with prior knowledge. The prior models are learned from training data and updated dynamically, based on segmentations of earlier images in the sequence. We demonstrate the potential of our approach in a hand gesture recognition application, where the combined use of segmentation and classification improves robustness in the presence of occlusion and background complexity.

1 Introduction

Image segmentation is one of the most basic yet most challenging problems of computer vision. Segmentation requires finding in an image semantically salient regions (or their bounding contours) associated with “objects”. Behavior classification in image sequences is an important higher level task towards comprehensive visual perception. By “behavior” of an object in an image sequence, we mean the temporal evolution of its attributes (such as position, orientation, shape, color, texture, etc.) apparent in the image sequence. The classification of object behavior refers to assigning one of several behavior class labels to each of its temporal evolution instances. For example, we would like to classify object motion (e.g., car turn directions at an intersection), classify motion and deformation (e.g., hand gestures, body motions), or classify intensity changes in a brain activation map for clinical purposes.

Conventionally, segmentation and behavior classification are solved separately and sequentially: one segments the image sequence, extracts the relevant features, and finally classifies their time evolution. However, the task of behavior classification can be facilitated if segmentation information is available. Reciprocally, segmentation can greatly benefit from considering the expected behavior

of the targeted object(s) (related to, e.g., its shape, color, texture), which is usually built into classification tasks as a priori models of the behavior classes to be distinguished. Therefore, benefits should accrue from a collaboration between image segmentation and behavior classification.

Our contribution in this paper is a coupled solution to image sequence segmentation and classification of object behavior, which enables the information related to each of them to enhance the results of both. To this end, we develop a new variational framework that smoothly integrates the two main sources of information: the target image sequence and the prior behavior models, which adapt dynamically to the latest segmented images through the classification strategy.

Variational methods underlie the mathematical formulation of numerous computer vision problems. The image segmentation problem has been formulated in terms of energy minimization, where one can seamlessly introduce various criteria describing the desired solution, such as smoothness, region homogeneity, edge correspondence, etc. Starting with the original active contour (snakes) model [1], continuing with the Mumford-Shah model [2], the introduction of the level set approach [3] and geodesic active contours [4–6], recent work has yielded versatile segmentation approaches such as [7, 8]. Statistical shape priors were introduced into active contours [9] and also into level set active contours [10–13] and the Mumford-Shah segmentation [14–16]. These techniques have made it possible to successfully segment a familiarly-shaped object in difficult cases. Variational methods for contour evolution have also been adopted for object tracking (e.g., [1, 17, 14]). Coherence between frames has been exploited by approaches based on Kalman filtering [18], particle filtering [19], and autoregressive models [20].

Our proposed framework deals simultaneously with the issues of image sequence segmentation and object behavior classification, leading to increased chances of success for both tasks through cooperation and information sharing. On the one hand, segmentation is improved by guidance towards the target object via probabilistic priors, offered by classification. On the other hand, classification is improved from the consideration of segmentation results captured from new images, while maintaining consistency with prior knowledge and with previous segmentations in the sequence. To our knowledge, the fusion of segmentation and behavior classification over image sequences is novel in the domain of variational image analysis, while it of course capitalizes on existing experience in the use of shape priors. The idea of combining segmentation and object recognition has previously yielded good results in the case of single images, both in variational [15] and non-variational [21–24] settings. Our work makes a significant contribution in that we address *image sequences* and the temporal problem of *object behavior classification*. To tackle this problem, we introduce a variational framework that incorporates dynamic probabilistic priors automatically obtained via a machine learning approach. We illustrate the potential of our proposed approach in a gesture recognition application, where the combination of segmentation and classification dramatically increases the tolerance to occlusion and background complexity present in the input image sequence.

We propose a *general framework* for the joint resolution of the two tasks—segmentation and behavior classification—whose components can be adapted

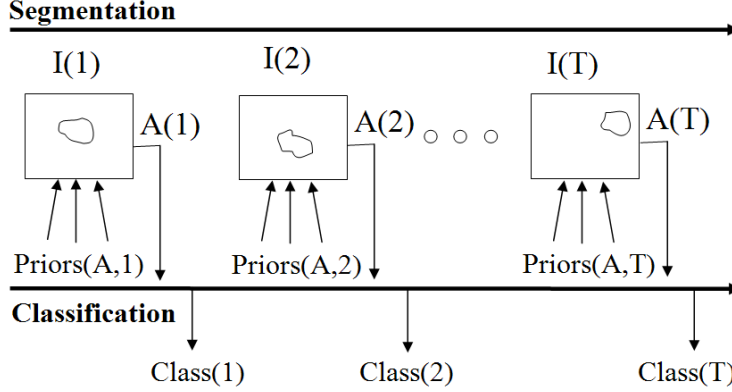


Fig. 1. Our approach: cooperation of segmentation and classification along the image sequence.

to suit the needs of a wide range of applications. The next section details the collaborating halves of our general framework, first behavior classification and then segmentation. A particular implementation of the framework is proposed in Section 3, which employs a specific image term and dynamic prior component, for the purposes of gesture recognition. Experimental results are presented at the end of Section 3. Section 4 concludes the paper.

2 Formulation of the Variational Framework

Our goal is to segment an image sequence and classify it in terms of object behavior. As illustrated in Fig. 1, the key idea in our framework is to interweave the classification and segmentation processes while iterating through the given image sequence. This enables them to collaborate in exploiting the available prior knowledge and to improve each other by sharing partial results obtained throughout the image sequence. More concretely, for each image in the sequence, classification offers dynamic probabilistic attribute priors to guide segmentation. These priors, which are based on training, adapt in time according to knowledge gained from new segmentations. In turn, segmentation detects, and supplies to classification, object attributes that best explain the image evidence consistently with the prior knowledge. These object attributes are used in the subsequent step of the classification, and so on, until the entire sequence is segmented and classified.

Note that by “attribute” we generically designate a visual property of the object of interest, representable as a functional $A(C, I)$ of the image I and of the object’s segmenting contour C (A is assumed to be differentiable with respect to C). The palette of such attributes is quite large, including all properties computable with boundary-based and/or region-based functionals, such as position, orientation, average intensity/color, or higher-order statistics describing texture.

2.1 Classification and its Cooperation with Segmentation

The behavior classification task aims at estimating, for a given time instance of an image sequence, the behavior class of the object, based on its observed attributes. Supposing for the moment that the attribute values are known, we need only find the generating behavior classes. We solve this problem using the machine learning concept of generative models [25], in particular Hidden Markov Models (HMMs) [26], where the observations are attribute values and the hidden states are the unknown behavior classes. Once trained on typical attribute evolution sequences, an HMM classifies new attribute sequences by estimating the most likely state sequence generating them.

We denote the states of the HMM (each corresponding to a behavior class) by $S = \{S_1, S_2, \dots, S_M\}$, the state at time t by q_t , and the attribute value at time t by $A(t)$. The HMM parameters are:

1. the initial state distribution $\pi = \{\pi_i\}$, with $\pi_i = P(q_1 = S_i)$, $i = 1..M$,
2. the state transition probability distribution $T = \{t_{ij}\}$, with $t_{ij} = P(q_{t+1} = S_j | q_t = S_i)$, $i, j = 1..M$, and
3. the state observation probability distributions (class likelihoods)

$$P(A(t) | q_t = S_i) = P_i(A(t)), \quad i = 1..M. \quad (1)$$

To support cooperation with the segmentation process, we require that these class likelihood functions $P_i(A(t))$ be differentiable with respect to $A(t)$.

Once having estimated the set λ of HMM parameters from training data, the HMM can be used to classify new attribute sequences. In order to assign a behavior class to each observation in a new sequence $A_{1..T} = \{A(1), A(2), \dots, A(T)\}$, we estimate the state sequence $q_{1..T}^{\text{opt}} = \{q_1, q_2, \dots, q_T\}^{\text{opt}}$ that best explains the observation sequence

$$q_{1..T}^{\text{opt}} = \arg \max_{q_{1..T}} P(q_{1..T} | A_{1..T}, \lambda) = \arg \max_{q_{1..T}} P(q_{1..T}, A_{1..T} | \lambda), \quad (2)$$

using the Viterbi algorithm [26]. At each time step t and for each state S_i , the Viterbi algorithm calculates the quantity

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_{1..t-1}, q_t = S_i, A_{1..t} | \lambda), \quad (3)$$

representing the highest probability along a state sequence, at time t , which explains the first t observations and ends in state S_i . This quantity is computed by initializing the δ s and then using the following recursion:

$$\delta_t(i) = (\max_j \delta_{t-1}(j) t_{ji}) P_i(A(t) | \lambda). \quad (4)$$

Finally, the optimal state sequence is determined by backtracking from these maximization results. Thus, the Viterbi algorithm iterates through the attribute sequence, computing its best estimate for the probability of different generating classes, given the knowledge accumulated in the HMM. We can use these estimates to guide the segmentation process. The idea is to run this algorithm

synchronously with the segmentation, using the attribute of the segmented object as the next observation as soon as it becomes available. Then, we incorporate the algorithm's best momentary class estimations as attribute priors for the segmentation of the next image in the sequence.

Now, suppose we have completed step $t - 1$ of both the segmentation and the Viterbi algorithm, so that attributes $A_{1..t-1}$ and $\delta_{t-1}(j)$, $j = 1..M$, are available. In order to segment $I(t)$, we use the maximum available *a priori* knowledge:

1. the predictions of each class i for the next attribute $A(t)$; i.e., the likelihood functions $P_i(A(t) | \lambda)$, $i = 1..M$
2. our relative confidence in the prediction of each class i , given by the Viterbi algorithm; i.e., the maximum probability of reaching state S_i at time step t , after having observed attributes $A_{1..t-1}$:

$$w_t(i) = \max_{j=1..M} \delta_{t-1}(j) t_{ji} = \max_{q_1, q_2, \dots, q_{t-1}} P(q_{1..t-1}, q_t = S_i, A_{1..t-1} | \lambda). \quad (5)$$

As prior information offered by each behavior class i , we shall use the product of these two quantities, which according to (4) is actually

$$\delta_t(A(t), i) = w_t(i) P_i(A(t) | \lambda), \quad i = 1..M; \quad (6)$$

i.e., δ_t as a function of the unknown attribute $A(t)$. Next, we explain how to introduce these class contributions into the segmentation framework.

2.2 Segmentation and its Cooperation with Classification

We take a variational approach to segmentation that incorporates the dynamic probabilistic priors offered by classification. For an image $I(t)$, these priors consist of the delta functions of the object attribute corresponding to each class i ; i.e., $\delta_t(A(t), i)$. We introduce these class contributions into the segmentation model by means of a competition mechanism, since we are searching for a single "winning" class that best accounts for the generation of the next observation.

To create a "competition" during segmentation among the priors associated with different classes, we employ a labeling mechanism motivated by [15]. For each prior i , we use one label L_i , a scalar variable that varies continuously between 0 and 1 during energy minimization and converges either to 0 or 1. The value of the set of labels $\mathcal{L} = (L_1, \dots, L_M)$ after convergence designates a "winner" among the attribute priors, corresponding to the probability which has been maximized through segmentation. Each of the prior terms carries a label factor equal to L_i^2 . Competition is enforced by constraining the label factors to sum up to 1 through the addition of the term $(1 - \sum_{i=1}^M L_i^2)^2$ to the segmentation energy.

Once having run our joint segmentation/classification framework on the first $t - 1$ frames of an image sequence, we segment $I(t)$ by minimizing with respect to the contour C and the labels \mathcal{L} the following energy functional:

$$E(C, \mathcal{L}, I(t)) = E_{\text{data}}(C, I(t)) + \alpha E_{\text{prior}}(C, \mathcal{L}, I(t)), \quad (7)$$

where α is a positive weight parameter. Here $E_{\text{data}}(C, I(t))$ can be any boundary-based or region-based segmentation energy, suitable to the application at hand (e.g., the energy proposed in [27]). The energy due to the priors is

$$E_{\text{prior}}(C, \mathcal{L}, I(t)) = - \sum_{i=1}^M \log(\delta_t(A(C, I(t)), i)) L_i^2 + \beta \left(1 - \sum_{i=1}^M L_i^2\right)^2, \quad (8)$$

where β is a positive constant and the δ function is defined in (6). The only assumptions regarding energy (8) are that the likelihood functions $P_i(A(C, I(t)))$ are differentiable with respect to the attribute $A(C, I(t))$ and the attribute is a differentiable functional of the contour C .

The minimization of (7) simultaneously with respect to the segmenting contour C and the label vector \mathcal{L} is performed via the calculus of variations and gradient descent. The contour C is driven by image forces due to $E_{\text{data}}(C)$, and by the M attribute priors due to $E_{\text{prior}}(C, \mathcal{L})$:

$$\frac{\partial C}{\partial \tau} = - \frac{\partial E_{\text{data}}(C, I(t))}{\partial C} - \alpha \frac{\partial E_{\text{prior}}(C, \mathcal{L}, I(t))}{\partial C}. \quad (9)$$

Here $\partial E_{\text{data}}(C, I(t))/\partial C$ can be derived through the calculus of variations for the particular chosen form of $E_{\text{data}}(C, I(t))$. The second term can be written as:

$$\begin{aligned} \frac{\partial E_{\text{prior}}(C, \mathcal{L}, I(t))}{\partial C} &= - \sum_{i=1}^M \frac{L_i^2}{\delta_t(A(C, I(t)), i)} \cdot \frac{\partial \delta_t(A(C, I(t)), i)}{\partial A} \cdot \frac{\partial A(C, I(t))}{\partial C}, \\ \text{with } \frac{\partial \delta_t(A(C, I(t)), i)}{\partial A} &= w_t(i) \frac{\partial P_i(A(C, I(t)) | \lambda)}{\partial A}. \end{aligned} \quad (10)$$

Derivatives $\partial P_i/\partial A$ and $\partial A(C, I(t))/\partial C$ are computed according to the particular likelihood function and attribute employed. In parallel with contour evolution, the labels compete to maximize the probability of the most likely prior given the image evidence:

$$\frac{\partial L_i}{\partial \tau} = \sum_{i=1}^M \delta_t(A(C, I(t)), i) L_i - \beta L_i \left(1 - \sum_{i=1}^M L_i^2\right). \quad (11)$$

The effect of these equations is that the label L_i corresponding to the maximum $\delta_t(A(C, I(t)), i)$ is driven towards 1 – i.e., the maximum δ_t is extremized – while the other labels are driven to 0.

In probabilistic terms, the minimization of our proposed energy using competing priors amounts to the maximization of the probability $\delta_t(A(t), i)$ with respect to both the attribute $A(t)$ and class i , subject to image-based constraints. Then, the segmentation of image $I(t)$ can be regarded as the joint estimation of the attribute value $A^*(t)$ and the class i^* as

$$\begin{aligned} (A^*(t), i^*) &= \arg \max_{A(t), i} \delta_t(A(t), i), \\ &\text{subject to image constraints } (A(t), I(t)). \end{aligned} \quad (12)$$

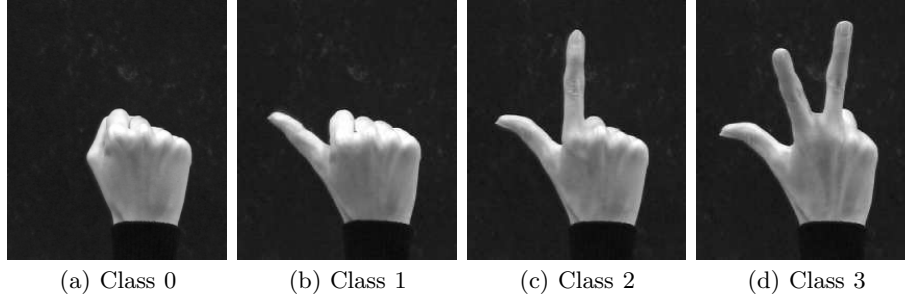


Fig. 2. Samples from the four gesture classes that we use in our application.

Thus, segmentation works concurrently towards the same goal as classification: maximizing the joint probability of the class and the observation at time t , while remaining consistent with previous observations, according to prior knowledge (through the HMM), and incorporating new information from image $I(t)$.

The segmentation of $I(t)$ yields $A(t)$, enabling the Viterbi algorithm to estimate $\delta_t(i)$ and $w_{t+1}(i)$, so that we can continue by segmenting $I(t+1)$ and repeat the cycle to the end of the image sequence. Finally, we obtain the classification of the image sequence as the most probable state sequence given the observations, by backtracking from the results of the Viterbi algorithm.

3 A Specific Implementation of our Framework for Hand Gesture Recognition

We now demonstrate the strength of our framework of Section 2 in a hand gesture recognition application. After describing the problem that we wish to address, we detail the particular implementation of our general framework, including the specific model that we use. Finally, we present the results obtained.

3.1 Application

In our application, we identify four gesture classes consisting of a right hand going through four finger configurations, exemplified in Fig. 2: fist (Class 0), thumb extended (Class 1), thumb and index finger extended (Class 2), and thumb, index, and middle finger extended (Class 3). Given an image sequence of such gestures, our goal is to perform joint segmentation and classification; i.e., for each image, extract the segmenting contour of the hand and determine the gesture class to which it belongs. Note that our gesture image sequence depicts finger-counting from 1 to 3 and back to 1, ending with the initial fist position; i.e., the following succession of gesture classes: 0,1,2,3,2,1,0. Our strategy is to train a 4-class HMM with such sequences and then to incorporate the HMM into our framework in order to segment and classify new sequences of this sort. In an application with more complicated gesture scenarios, the optimal number of gesture sub-states could be determined as in [28].

3.2 Solution using the proposed framework

For this application, the object attribute employed within our framework is the contour segmenting the hand $A(C, I) = C$. Using the level set approach [3], we represent the contour by the level set function (LSF) $\phi : \Omega \rightarrow \mathbb{R}$, chosen to be the signed distance function to the contour, so that $C \equiv \{(x, y) : \phi(x, y) = 0\}$.

Regarding the segmentation energy (7), as a data term we use the piecewise constant Mumford-Shah model as in [27]:

$$E_{\text{data}}(\phi) = \iint_{\Omega} (I - \mu_+)^2 H(\phi) dx dy + \iint_{\Omega} (I - \mu_-)^2 (1 - H(\phi)) dx dy + \nu \iint_{\Omega} |\nabla H(\phi)| dx dy, \quad (13)$$

where H is the Heaviside function and μ_+ , μ_- are mean image intensities over the positive, respectively negative regions of ϕ . The prior energy is given by

$$E_{\text{prior}}(\phi, \mathcal{L}) = - \sum_{i=1}^M \log(\delta_t(\phi, i)) L_i^2 + \beta \left(1 - \sum_{i=1}^M L_i^2\right)^2, \quad (14)$$

where $\delta_t(\phi, i) = w_t(i) P_i(\phi)$. We use a local Gaussian probability model for each class i :

$$p_i^{(x,y)}(\phi) = \frac{1}{\sqrt{2\pi}\sigma_i((x,y))} e^{-\frac{(\phi(x,y) - \phi_i(x,y))^2}{2\sigma_i^2(x,y)}}, \quad (15)$$

where $(x, y) \in \Omega$ is an image location, ϕ_i is the average LSF of class i , and the variance $\sigma_i(x, y)$ models the local variability of the level set at (x, y) . Assuming densities independent across pixels, the likelihood of an LSF ϕ , offered by class i , is the product of these densities over the image domain:

$$P_i(\phi) = \prod_{(x,y) \in \Omega} p_i^{(x,y)}(\phi). \quad (16)$$

Substituting likelihoods $P_i(\phi)$ in (14) and augmenting by similarity transformations (including translation, rotation, and scale) that align each prior i with contour ϕ , the prior energy becomes:

$$E_{\text{prior}}(\phi, \mathcal{L}, \boldsymbol{\tau}^{i=1..M}) = \sum_{i=1}^M \left(-\log w_t(i) + \iint_{\Omega} \left(\frac{(\phi(x, y) - \phi_i(h_{\boldsymbol{\tau}^i}(x, y)))^2}{2\sigma_i^2(h_{\boldsymbol{\tau}^i}(x, y))} + \log \sigma_i(h_{\boldsymbol{\tau}^i}(x, y)) \right) dx dy \right) L_i^2 + \beta \left(1 - \sum_{i=1}^M L_i^2\right)^2. \quad (17)$$

Here, $\boldsymbol{\tau} = \{s, \theta, T_x, T_y\}$ are the parameters of a similarity transformation

$$h_{\boldsymbol{\tau}}([x \ y]^T) = s \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix}. \quad (18)$$

Parameters $\boldsymbol{\tau}^i$ for each class i evolve during segmentation according to their corresponding gradient descent equations, leading to the minimization of (17).

3.3 Training the model

In the training phase, we estimated the parameters of the HMM (see, e.g., [26]) using a labeled, rigidly aligned sequence of LSFs corresponding to a manual segmentation of the mentioned gesture sequence (0,1,2,3,2,1,0). We used the method in [13] to obtain smooth estimates of the mean ϕ_i and variance σ_i for each gesture class i .

3.4 Results

In the testing phase, we run classification and segmentation jointly on new image sequences of a hand performing the same succession of gestures in front of a complex background, in the presence of occlusions and noise. By virtue of the prior information supplied by classification, segmentation is able to cope with severe occlusions, as can be seen in Fig. 3(a)–(d), (i)–(l). Figure 3(e)–(h), (m)–(p) shows that conventional segmentation of the same sequences is clearly inferior, failing to recover the desired shape of the object because of the occlusions.

Fig. 4 shows the classification results for the first test sequence, which correctly follow our understanding of the executed gestures. Moreover, the final classification yielded by the Viterbi algorithm corresponds to the partial classification results used to guide segmentation throughout the sequence. This can be seen in Fig. 4, which exhibits, as functions of time (frame), (a) the final classification, (b) the delta functions of each class, and (c) the prior confidence of each class (the w function) used as input to the segmentation. The w values are scaled with respect to their maximum value for every frame.

Even though our chosen test application might not seem especially challenging, the proposed framework can potentially be applied to much more complicated scenarios. Its power lies in its flexibility: it allows a large variety of implementations, capitalizing on existing expertise in both probabilistic learning and variational segmentation.

4 Conclusion

We have introduced a novel variational framework for the simultaneous segmentation and object behavior classification of image sequences. Cooperation between the segmentation and classification processes facilitates a mutual exchange of information, which is beneficial to their joint success. In particular, we employed a classification strategy based on generative models that provides dynamic probabilistic attribute priors to guide image segmentation. These priors allow the segmentation process to work towards the same goal as classification, by outlining the object that best accounts for both image data and prior knowledge encapsulated in the generative model. We illustrated the potential of our general framework in a hand gesture analysis application, where we successfully segmented and classified image sequences of a gesturing hand before a complex background, in the presence of occlusions and noise. Future directions of our work will include the use of more complex attribute priors that would be better suited to challenging, under-constrained problems in high-dimensional spaces, such as the inference of 3D hand pose and behavior from monocular images.

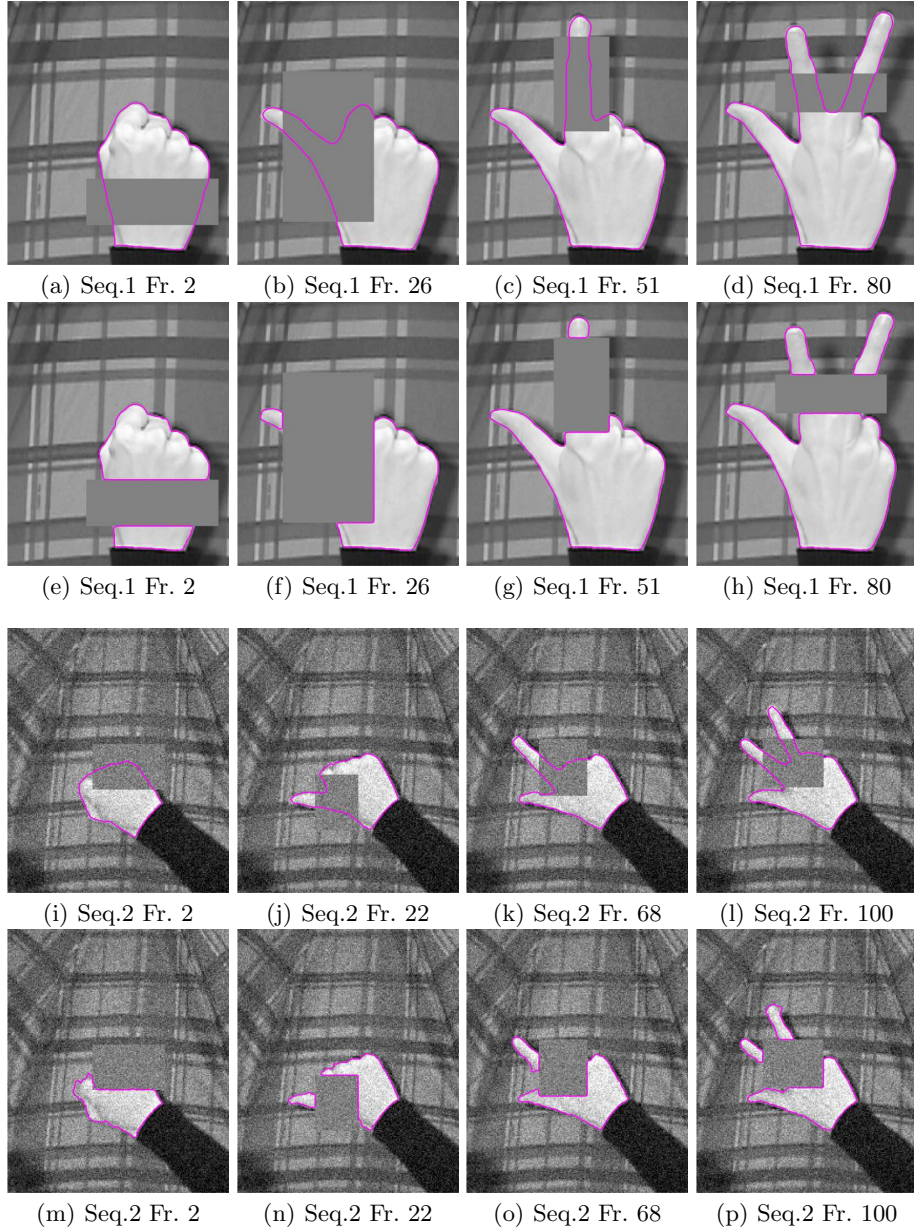


Fig. 3. (a)–(d), (i)–(l) Segmentation with the proposed framework of two image sequences in the presence of occlusion, background complexity and noise (second sequence). (e)–(h), (m)–(p) Conventional segmentation of the same image sequences.

5 Acknowledgements

The collaborative work between the authors of this paper has been made possible via the PAI De Stael 08392QA French-Swiss program. The authors would like to thank Xavier Bresson for fruitful discussions supportive of this paper.

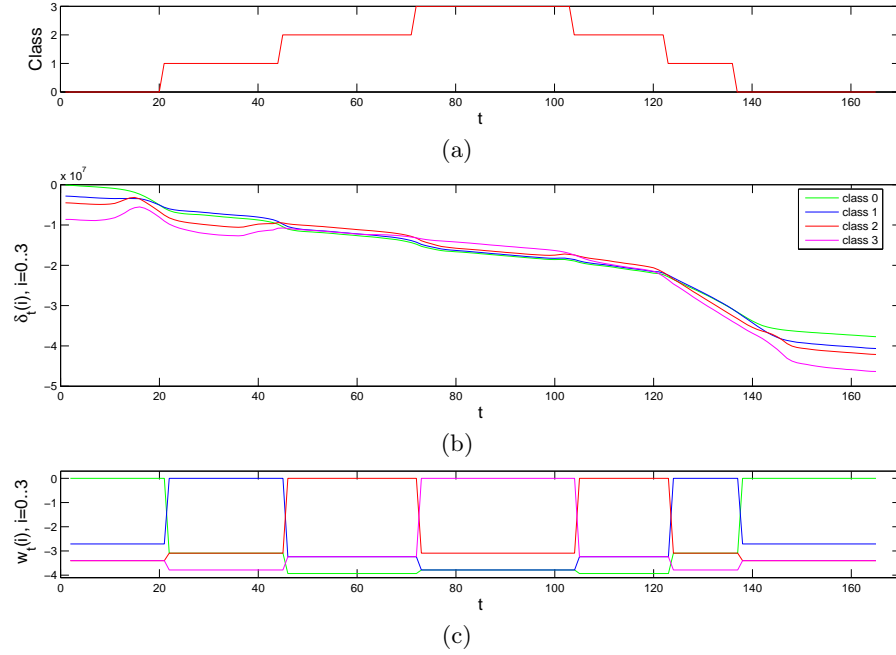


Fig. 4. Classification results plotted per frame. (a) Final classification. (b) Delta functions of each class. (c) Prior confidence of each class used as input to the segmentation.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* **1** (1987) 321–331
2. Mumford, D., J.Shah: Optimal approximations by piecewise smooth functions and associated variational problems. *Communications in Pure and Applied Mathematics* **42** (1989) 577–685
3. Osher, S., Sethian, J.: Fronts propagating with curvature-dependent speed: Algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics* **79** (1988) 12–49
4. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. In: *Proc. IEEE Intl. Conf. on Comp. Vis., Boston, USA* (1995) 694–699
5. Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., Yezzi, A.: Gradient flows and geometric active contour models. In: *Proc. IEEE Intl. Conf. on Comp. Vis.* (1995) 810–815
6. Malladi, R., Sethian, J., Vemuri, B.: Shape modeling with front propagation: A level set approach. *IEEE PAMI* **17** (1995) 158–175
7. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision* **50(3)** (2002) 271–293
8. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision* **46(3)** (2002) 223–247

9. Cootes, T., Beeston, C., Edwards, G., Taylor, C.: Unified framework for atlas matching using active appearance models. *Intl Conf. Inf. Proc. in Med. Imaging* (1999) 322–333
10. Leventon, M., Grimson, W., Faugeras, O.: Statistical shape influence in geodesic active contours. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. (2000) 316–323
11. Tsai, A., Yezzi, A., Wells, W., Tempny, C., Tucker, D., Fan, A., Grimson, W., Willsky, A.: Model-based curve evolution technique for image segmentation. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. (2001) 463–468
12. Chen, Y., Tagare, H., Thiruvankadam, S., Huang, F., Wilson, D., Gopinath, K., Briggs, R., Geiser, E.: Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision* **50(3)** (2002) 315–328
13. Paragios, N., Rousson, M.: Shape priors for level set representations. In: *European Conference in Computer Vision*. Volume 2. (2002) 78–92
14. Cremers, D., Osher, S., Soatto, S.: Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: Teaching level sets to walk. *Pattern Recognition* **3175** (2004) 36–44
15. Cremers, D., Sochen, N., Schnör, C.: Multiphase dynamic labeling for variational recognition-driven image segmentation. In: *European Conference on Computer Vision*. Volume 3024. (2004) 74–86
16. Bresson, X., Vandergheynst, P., Thiran, J.P.: A variational model for object segmentation using boundary information and shape prior driven by the Mumford-Shah functional. *International Journal of Computer Vision* **28(2)** (2006) 145 – 162
17. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for motion estimation and tracking. *Computer Vision and Image Understanding* **97** (2005) 259–282
18. Terzopoulos, D., Szeliski, R.: Tracking with Kalman snakes. *Active vision* (1993) 3–20
19. Rath, Y., Vaswani, N., Tannenbaum, A., Yezzi, A.: Particle filtering for geometric active contours with application to tracking moving and deforming objects. In: *Proc. CVPR*. Volume 2. (2005) 2–9
20. Cremers, D., Funka-Lea, G.: Dynamical statistical shape priors for level set based tracking. In *LNCS*, S., ed.: 3rd. Workshop on Variational, Geometric and Level Set Methods in Computer Vision. Volume 3752. (2005) 210–221
21. Tu, Z., Chen, X., Yuille, A., Zhu, S.: Image parsing: Segmentation, detection, and recognition. In: *ICCV*. (2003) 18–25
22. Leibe, B., Leonardis, A., Schiele, B.: Combined object categorization and segmentation with an implicit shape model. In: *ECCV Workshop on SLCV*. (2004)
23. Ferrari, V., Tuytelaars, T., Gool, L.V.: Simultaneous object recognition and segmentation by image exploration. In: *ECCV*. (2004)
24. Kokkinos, I., Maragos, P.: An Expectation Maximization approach to the synergy between image segmentation and object categorization. In: *ICCV*. (2005) 617–624
25. Bishop, C.: *Pattern Recognition and Machine Learning*. Springer (2006)
26. Rabiner, L.: A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE* **77(2)** (1989)
27. Chan, T., Vese, L.: Active contours without edges. *IEEE Transactions on Image Processing* **10(2)** (2001) 266–277
28. Stenger, B., Ramesh, V., Paragios, N., Coetzee, F., Bouhman, J.: Image parsing: Segmentation, detection, and recognition. In: *ICCV*. (2003) 18–25